

Réseaux bayésiens et inférences causales : quels enseignements pour l'indicateur synthétique de la qualité des systèmes éducatifs des pays de l'OCDE ?¹

François-Marie Gerard

Bernard Hugonnier

Sacha Varin

Résumé

La recherche en éducation est largement basée sur des études observationnelles. La possibilité de démontrer des relations causales dans de telles études est en débat. Cependant, plusieurs méthodes d'analyse causale pour ces données ont été développées ces vingt dernières années. La présente recherche vise à identifier les relations causales entre les six critères définissant l'indicateur synthétique de la qualité des systèmes éducatifs (ISQ) et son score final en 2018. Pour cela sont utilisés les réseaux bayésiens causaux et plus particulièrement les graphes acycliques orientés qui permettent d'identifier les causalités.

Mots-clés : réseaux bayésiens, indicateur synthétique de la qualité des systèmes éducatifs « ISQ », graphe acyclique orienté, causalité

Abstract

Educational research is largely based on observational studies; but the possibility of demonstrating causal relationships in such studies remained hypothetical for a very long time. However, several methods of causal analysis for these data have been developed over the last twenty years, including Bayesian networks and more particularly directed acyclic graphs which are used to analyse the causal relationships among the criteria of the quality indicator of the OECD countries. The purpose of this research is to use these methodological advances for the criteria of the most recent synthetic quality indicator for OECD countries (SQI).

Keywords : bayesian networks, synthetic indicator "ISQ", directed acyclic graph, causality

¹ Publié en anglais, référence officielle : GERARD, F.-M., HUGONNIER, B. & VARIN, S. (2023). Bayesian networks and causal discovery: what lessons for the synthetic indicator of the quality of education systems in OECD countries?, *International Journal of Research in Social Sciences*, Volume 13 Issue 1, January 2023, ISSN: 2249-2496, THOMSON REUTERS RESEARCHER ID: L-5180-2015, 73-81.

1. Introduction

Selon Talbot (2012), la manière la plus simple et la plus sûre de mettre en évidence des inférences causales nécessite des expériences randomisées, très utilisées en médecine par exemple. On divise alors les unités expérimentales aléatoirement en deux groupes qui reçoivent, sans le savoir, soit le traitement qu'on veut tester, soit un autre traitement ou un placebo (recherche dite en double aveugle). La différence des effets dans les deux groupes pourra ainsi prouver l'efficacité du traitement.

Qu'en-est-il lorsqu'une étude randomisée est impossible ? Doit-on abandonner complètement l'idée de rechercher des inférences causales ? La réponse donnée par Judea Pearl (2000) est un « non » catégorique. Mais cela se fera à un certain prix : satisfaire des hypothèses et d'autres conditions.

L'approche proposée par Pearl consiste à tracer un graphique acyclique orienté (*Directed acyclic graph* ou DAG). Il s'agit d'un graphique où les variables sous examen sont reliées par des flèches qui indiquent le sens et l'importance des causalités (Pearl, 1995, 2000, 2003, 2009).

Dans la présente recherche, en utilisant cette méthode, nous avons cherché à mettre en évidence les causalités au sein de l'indicateur synthétique de la qualité des systèmes éducatifs des pays de l'OCDE (ISQ) le plus récent, soit celui établi à partir des données PISA de 2018. Cet indicateur est composé de six critères : l'efficacité, l'efficacité, l'équité, l'engagement des parents, l'engagement des élèves et l'engagement des enseignants. Chacun de ces critères est défini par un score, la moyenne de l'ensemble constituant le score final de l'ISQ. Celui-ci est calculé de manière diachronique (selon le rythme des études PISA qui fournissent la plupart des données primaires). Nous ne détaillons pas davantage ici les six critères et l'ISQ ; les lecteurs intéressés pourront en prendre connaissance dans nos écrits précédents : Gerard, Hugonnier et Varin, 2017, 2018, à paraître.

Les deux principaux objectifs visés par cette recherche sont

- d'une part – grâce au graphe acyclique orienté – d'identifier, de mesurer, de comprendre et d'interpréter toutes les relations causales entre ces différents critères ;
- d'autre part de mesurer la force des six relations fonctionnelles entre chacun des critères et le score final ISQ 2018, ce dernier n'étant que la moyenne des scores obtenus par chaque critère.

Les résultats permettront de déterminer la confiance qu'on peut avoir dans la force et la direction de chacune des flèches du DAG.

Dans cet article, nous exposons d'abord le cadre théorique en lien avec la recherche de causalités ; nous présentons ensuite les outils méthodologiques utilisés pour mettre en évidence les différents effets causaux directs. Enfin, nous exposons les principaux résultats et les discutons avant d'en tirer les conclusions.

2. Cadre théorique

Avant de débiter cette partie, nous tenons à préciser que deux tâches principales peuvent être distinguées dans le domaine de la causalité : la découverte causale et l'inférence causale. Dans cette dernière, on sait, grâce à des données expérimentales, que tel(s) lien(s) causal(aux) existe(nt) entre telles variables. Les relations causales y sont donc définies dès les début de l'analyse. Notre travail s'inscrit par contre clairement dans le domaine de la découverte causale, à savoir, à partir d'un ensemble de données d'observation, essayer de déduire toutes les relations causales. La découverte causale ne suppose aucune relation a priori entre les variables impliquées. C'est le processus de découverte qui permet de déduire les relations directement des variables.

Dans cette partie, à la suite d'une revue de littérature concernant la modélisation et l'analyse des réseaux causaux depuis le début des années 2000, nous abordons la notion de réseaux bayésiens par une présentation de la théorie des graphes et la méthode, issue des réseaux bayésiens causaux, appelée graphe acyclique orienté (ou DAG en anglais) ; méthode qui permet d'identifier les causalités des données observationnelles (Maathuis, Kalisch, Bühlmann, 2009). Enfin, nous esquissons une présentation de la notion de réseaux bayésiens causaux.

2.1 Revue de la littérature de la modélisation des réseaux causaux et leur analyse

L'inférence causale et les différentes théories qui y sont associées, ont été largement abordées et résumées dans de nombreux ouvrages (Pearl, 2000, 2003, 2009 ; Spirtes et al., 2000 ; Morgan and Winship, 2014 ; Imbens and Rubin, 2015 ; Vanderweele, 2015 ; Peters and Janzing, 2017 ; Hernàn and Robins, 2018). Les approches de modélisation et d'analyse des réseaux causaux ont émergé dans les dernières décennies du 20^e siècle et leur utilisation est actuellement en plein essor.

Les modèles graphiques probabilistes, et plus particulièrement les réseaux bayésiens ont été primitivement développés dans les années 1980 par Judea Pearl qui a très largement contribué à la théorie des modèles structuraux permettant l'identification et l'estimation des effets causaux à partir de données observationnelles. C'est également lui qui a développé le graphe DAG (*directed acyclic graph*) qui fait partie des réseaux bayésiens causaux.

2.2 Les réseaux bayésiens

Les réseaux bayésiens sont des modèles graphiques probabilistes qui représentent des variables aléatoires et leurs dépendances conditionnelles ainsi que des tables de probabilités permettant d’acquérir, de capitaliser et d’exploiter des connaissances (Chickering, Geiger et Heckerman, 1995 ; Chickering et Heckerman, 1996 ; Naïm, Wuillemin, Leray, Pourret et Becker, 2007).

Ils permettent de représenter des probabilités et de calculer efficacement celles utiles à la prise de décisions. L’intérêt particulier des réseaux bayésiens est de tenir compte simultanément de connaissances a priori d’experts du sujet et de l’information contenue dans les données.

Étant donné la grande flexibilité des réseaux bayésiens, ils ont été utilisés dans de nombreuses disciplines : finance, économie, médecine, robotique, génie civil, géologie, génétique, criminologie, écologie, industrie, etc. (Ben Hassen, Masmoudi et Rebai, 2008 ; Lauritzen, 1996 ; Naïm, Pourret et Marcot, 2008).

2.3 Brève présentation de la théorie des graphes

Les graphes constituent une méthode de pensée qui permet de modéliser une grande variété de problèmes en se ramenant à l’étude de sommets et d’arêtes. Les sommets sont souvent représentés par des variables et les arêtes peuvent être dirigées – dans ce cas il y a présence d’une flèche indiquant la direction, c’est le cas des DAG (Directed Acyclic Graph) – ou non dirigées, c’est le cas par exemple des champs de Markov (Pearl, 1988).

Plus précisément, un graphe $G = (V, E)$ consiste en un ensemble de représentations graphiques de nœuds (V) et d’arêtes (E) qui les connectent. Les nœuds représentent des variables aléatoires $V = (V_1, \dots, V_p)$ et les arêtes les liens entre les variables.

Dans notre étude, c’est le graphe acyclique dirigé (DAG) qui nous intéresse. Il se distingue par la présence de flèches dirigées unidirectionnelles et aucune boucle. Les Figures 1 et 2 illustrent la différence entre les graphes dirigés cyclique et acycliques.

Figure 1 : Graphe cyclique dirigé : présence d’une boucle entre X1, X2 et X3.

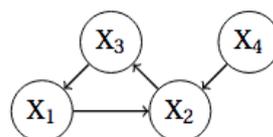
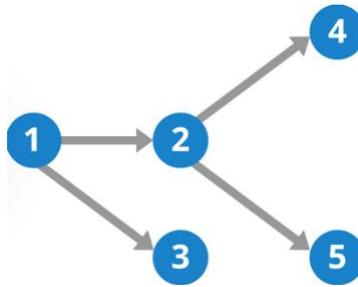


Figure 2 : Graphe acyclique dirigé (DAG) : présence de flèches dirigées avec une seule direction et aucune boucle.



2.4 Les réseaux bayésiens causaux

On distingue deux sortes de réseaux : ceux où les flèches ne doivent pas forcément être interprétées en termes de causalité ; et ceux où les flèches d'un graphe orienté, représentent bien des relations de cause à effet directes entre l'ensemble des variables. Les réseaux bayésiens causaux sont en fait une extension des réseaux bayésiens classiques où toute relation entre variables correspond à une relation causale (Pearl, 2000). Un réseau bayésien causal est un réseau bayésien, mais avec une propriété principale supplémentaire (Pearl, 2000) qui réside dans le fait que chaque ensemble de flèches parent \rightarrow enfant ($Pa(V_i) \rightarrow V_i$) ne représente plus seulement une dépendance probabiliste, mais une relation causale. Dans un réseau bayésien causal, chaque table de probabilités conditionnelles (CPD) représente un processus stochastique dans lequel les valeurs de V_i sont choisies en fonction des valeurs de $Pa(V_i)$ mais pas l'inverse. Cette propriété fait que les réseaux bayésiens causaux sont des modèles graphiques probabilistes dont la structure est encore plus lisible par des experts du domaine considéré. Un réseau bayésien causal permet ainsi de vérifier des hypothèses de causalité.

3. Méthodologie

Dans cette partie, nous nous intéressons aux notions très importantes de force/présence et direction des flèches en abordant la construction du DAG grâce à l'algorithme *hill climbing* et la technique du *bootstrap* utilisant le logiciel R (R Core Team, 2022) et plus particulièrement les packages *bnlearn* (Scutari, 2017) et *pcalg* (Kalisch, Mächler, Colombo, Maathuis, Bühlmann, 2012 ; Maathuis, Colombo, Kalisch, Bühlmann, 2010). Nous discutons aussi les hypothèses/conditions devant être remplies pour que le DAG puisse être interprété en termes de causalité.

3.1 Mesurer la force/présence et la direction des flèches

Mesurer le degré de confiance des flèches dans un réseau bayésien tel que le graphique DAG relève d'un problème majeur en inférence causale. Friedman, Goldszmidt et Wyner (1999) ont introduit un moyen de quantifier ce niveau de confiance. Il s'agit de générer plusieurs DAG par

bootstrap non paramétriques et d'estimer la fréquence d'apparition des flèches ; un bootstrap consistant en une réplique élevée des données selon la technique de l'échantillonnage.

Nous utilisons la fonction `boot.strength()` du package *bnlearn* (Scutari, 2017). Cette fonction estime la force/présence de chaque flèche selon sa fréquence empirique dans un ensemble de DAG construits à partir d'échantillons *bootstrap*. Elle calcule la probabilité de chaque flèche (*modulo* sa direction) et les probabilités des directions de chacune à condition que celle-ci soit présente dans le DAG.

En d'autres termes, la force et la direction de la relation sont mesurées et définies par la fréquence d'apparition des flèches dans les DAG construits par *bootstrap*. L'objectif est de découvrir dans quelles proportions la présence et la direction d'une flèche entre deux variables apparaissent dans les 100 000 DAG construits par *bootstrap*.

3.2 L'algorithme *hill climbing*

Avant de présenter l'algorithme qui nous a permis d'effectuer ces opérations, il est primordial d'expliquer la raison pour laquelle nous avons retenu l'algorithme *hill climbing* plutôt qu'un autre : comparé à d'autres algorithmes, c'est celui qui minimise le score BIC (critère/score d'information bayésien, voir en 3.2.1) lors d'une analyse de cross-validation en utilisant la méthode *hold-out*.

L'algorithme *hill climbing* (Gámez, Mateo et Puerta, 2011) appelé également *greedy search* fonctionne sur la maximisation d'un score de réseau traduisant sa bonne adéquation avec les données disponibles, score que l'algorithme essaie d'optimiser. Cet algorithme se concentre sur la construction d'un DAG de manière globale, le score étant calculé sur l'ensemble de la structure. Ce score permet d'estimer la qualité d'un réseau dans son intégralité au vu des observations. Le problème de construction se transforme alors en un problème d'optimisation de ce score. Chacun des scores doit maximiser la probabilité $P(G|D)$ du graphe G sachant les observations. Or, d'après la formule de Bayes : $P(G|D) = \frac{P(D|G) \times P(G)}{P(D)}$, la probabilité $P(D|G)$ représente la vraisemblance marginale des données sachant le modèle, celle-ci doit être maximale.

Le grand problème de cet algorithme est que, s'il peut mener à un score maximum, celui-ci peut correspondre en fait à un maximum local (et non global). Pour contrer ce problème, la meilleure solution est alors d'effectuer plusieurs fois l'algorithme (*iterated hill climbing*) et de retenir le graphe ayant le meilleur score. Le score que nous utilisons pour l'algorithme est le BIC. Le *hill climbing*, pour construire le DAG, va donc maximiser la vraisemblance marginale grâce au score BIC.

3.2.1 Le score BIC

De nombreux scores existent tels que l'AIC, le BIC, le BD, etc. Le BIC est l'un des scores les plus souvent utilisés, il permet l'utilisation d'informations *a priori* et se présente sous la forme suivante : $BIC = \log(P(D|G)) - \frac{d}{2} \log(n)$ où le terme de droite est un terme de pénalité, limitant les structures trop compliquées et le sur-ajustement, et où « n » représente le nombre d'observations dans l'échantillon et « d », le nombre de paramètres associés au réseau qui se calcule grâce au nombre de variables dans le réseau. La première partie représente la probabilité d'avoir ces données D sachant le graphe G . L'objectif étant donc de maximiser le score BIC. Le principal avantage du score BIC est – et c'est pour cette raison que nous l'avons retenu – que la taille de l'échantillon n'a pas besoin d'être grande pour que les résultats soient valides. Plus précisément, une taille d'échantillon entre 20 et 30 suffit largement (Murphy, 2007).

3.3 Construction du DAG par la technique du *bootstrap*

Au vu de la taille relativement restreinte de notre échantillon, le DAG est construit grâce à la technique du *bootstrap* avec 100 000 répliques. Ce nombre étant très élevé, nous obtenons une estimation des fréquences empiriques de la force/présence et de la direction des flèches très précises. Au final, le DAG est obtenu en ne conservant que les flèches qui sont présentes dans au moins 70% des DAG *bootstrap*. Il n'existe pas de valeur limite ; en règle générale les chercheurs interprètent comme relation causale les proportions supérieures à 80% pour la présence de flèches et supérieures à 50% pour leur direction. Il s'agit d'un choix subjectif. Pour retenir la présence d'une flèche, nous nous sommes fixés un seuil minimal de 70%.

3.4 Les conditions à satisfaire

Afin de pouvoir interpréter les flèches en tant que causalités, plusieurs conditions doivent être satisfaites.

La première condition – il s'agit de l'hypothèse principale faite par les réseaux bayésiens causaux – est appelée condition de Markov causale. Un DAG doit vérifier cette condition. Deux variables non causalement reliées directement sont indépendantes conditionnellement à leur set de parents markoviens communs. En d'autres termes, chaque variable est indépendante de ses non descendantes conditionnellement à ses parents. Cette hypothèse permet de faire la distinction entre corrélation et causalité. L'objet principal de la condition de Markov causale est de dire que toute la connaissance statistique, nécessaire pour la modélisation du processus en cours, est contenu dans le présent (pour limiter le nombre de variables). L'avantage de cette

hypothèse est qu'elle peut presque toujours être staisfaite en pratique, il suffit effectivement de donner la bonne définition du « présent ». Par exemple, puisque nous basons notre étude sur les données de 2018, nous dirons que le présent est l'année 2018.

La deuxième condition est l'hypothèse de fidélité (*faithfulness*) entre le graphe et la distribution de probabilité sous-jacente à nos données « P ». Il s'agit de l'existence d'un réseau bayésien qui soit la P -map du modèle d'indépendance associé à la distribution de probabilité « P ». En d'autres termes, un modèle est fidèle s'il n'oublie aucune indépendance conditionnelle.

La troisième condition est la suffisance causale. Suivant cette condition, toutes les variables potentiellement causales sont prises en compte dans l'analyse. L'ensemble des variables du DAG est suffisant pour représenter toutes les relations d'indépendances conditionnelles qui pourraient être extraites des données. Pouvons-nous prétendre à la non-présence de variables confondantes non mesurées ? Une analyse grâce à l'algorithme *Fast Causal Inference* (FCI) (Spirtes et al., 1999) révèle que nous n'avons pas assez d'information pour répondre à cette question. Cela signifie que nous pouvons n'avoir aucune variable confondante ou alors être en présence de variables confondantes. Il serait surprenant qu'il n'y ait pas de variables confondantes non mesurées dans une thématique aussi complexe qu'est la qualité d'un système éducatif, c'est la raison pour laquelle le DAG présenté ici est une solution potentielle, mais il se peut qu'il y ait d'autres DAG qui contiennent des variables confondantes non mesurées.

Ces trois conditions doivent être satisfaites afin de pouvoir interpréter les DAG en termes de causalité. Il est primordial de souligner que ces trois conditions ne sont pas des hypothèses que l'on peut tester/vérifier aisément : les données ne peuvent pas nous dire si ces trois hypothèses sont appropriées. La question de savoir si elles sont satisfaisantes relève de connaissances et de jugement spécifiques. Seuls des chercheurs expérimentés connaissant très bien le sujet et les variables du DAG peuvent déterminer si les trois conditions sont satisfaites. Le slogan de N. Cartwright (1994) est à cet égard très évocateur. *No causes in, no causes out*. Si on veut être en mesure d'estimer un effet causal direct, nous avons besoin d'une compréhension qualitative générale de la structure causale dans laquelle cet effet est intégré.

Il est également primordial de parler d'une autre condition tout aussi importante avant de pouvoir interpréter les DAG de manière causale : les variables ne doivent pas être fonctionnellement dépendantes, car dans ce cas elles peuvent créer un biais dans le DAG. Nos données sont fonctionnellement dépendantes : en effet, la variable « score final ISQ 2018 » est la moyenne des six autres variables. Notons que pour la validité du DAG, nous avons retiré la variable « Score final ISQ 2018 » de l'analyse afin d'éviter les dépendances fonctionnelles. Il est ressorti trois flèches de l'analyse. Par la suite, nous avons rajouté la variable « Score final

ISQ 2018 » dans le modèle comme dépendante fonctionnellement des six autres variables. En outre, il est important pour notre recherche de mesurer la force et la direction des dépendances fonctionnelles, car elles amènent à une meilleure compréhension de la confiance qu'on peut avoir dans chaque critère qui compose notre indicateur synthétique (ISQ). Les forces des six variables qui constituent la variable moyenne « Score final ISQ 2018 » ne sont pas égales, car il existe une légère colinéarité ($VIF < 2$) dans nos données, ce qui fait que les variables ne sont pas toujours reliées à « Score final ISQ 2018 » en même temps.

Afin de distinguer les dépendances probabilistes de celles qui sont fonctionnelles dans le DAG, nous présentons des flèches pleines pour les dépendances probabilistes et des flèches en traitillés pour les dépendances fonctionnelles.

Compte tenu du fait que le jeu de données utilisé pour construire le DAG est de taille limitée, nous avons opté pour des approches de type *bagging/bootstrap* afin d'obtenir des résultats plus robustes.

Une dernière condition très importante à relever, en rapport direct avec l'algorithme retenu pour construire le DAG : ne pas tomber dans un optimum local lors de l'utilisation de l'algorithme *hill climbing*. Un optimum local ne permettrait pas d'avoir le DAG le plus fiable. Afin de contrer cet optimum local, nous avons utilisé des itérations consistant à répéter 100 000 fois l'algorithme. Relevons que l'algorithme *hill climbing* tombe forcément dans un optimum local, on ne peut en effet jamais être certain de la globalité de la solution courante, même après itérations. Il faut s'y faire et il faut également se faire à l'idée que le paysage des réseaux bayésiens n'est pas très calme, car y prévalent nombre d'optimums locaux. Pour toutes ces raisons, nous devons être prudents dans l'interprétation causale de notre DAG : ainsi, nous employons le conditionnel et utilisons le DAG comme outil permettant de fournir des pistes d'interprétations qui devront être vérifiées, si possible, expérimentalement.

4. Résultats

Dans cette partie, nous présentons tous les résultats issus du DAG (figure 3), y compris les dépendances fonctionnelles (indiquées en traitillés).

4.1 Présence/force et direction des flèches

Comme indiqué dans la méthodologie, nous n'avons conservé que les flèches qui ont une force/présence d'au moins 70%, en symbolisant celle-ci par l'épaisseur de la flèche.

Dans le tableau 1 figurent les deux proportions (présence et direction) pour les neuf flèches du DAG.

Tableau 1 : Présence et direction des flèches

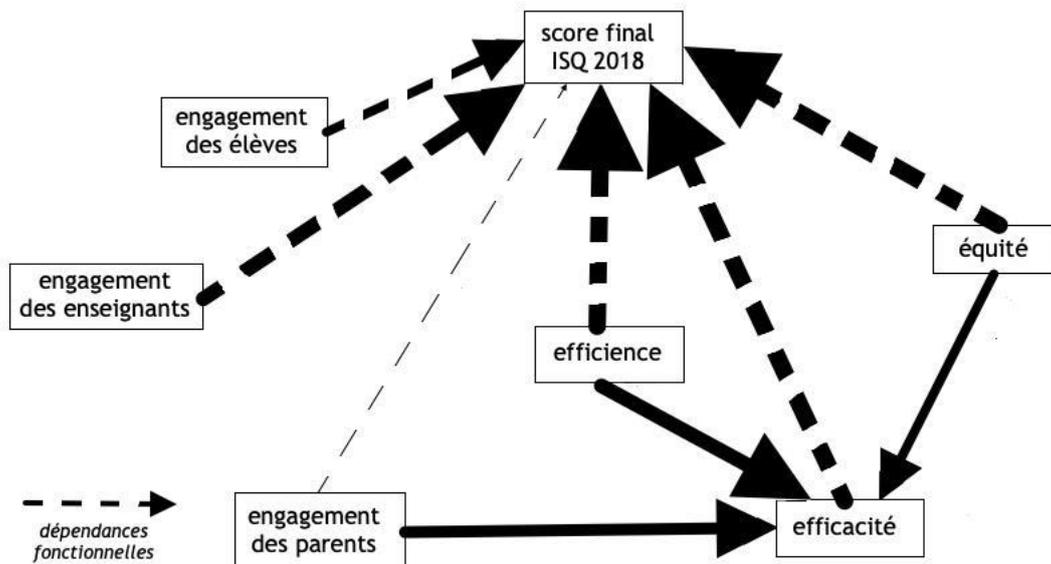
De	À	Présence	Direction
Engagement des élèves	Score final ISQ 2018	83%	91%
Engagement des enseignants	Score final ISQ 2018	95%	76%
Engagement des parents	Score final ISQ 2018	70%	91%
Efficacité	Score final ISQ 2018	98%	51%
Efficiace	Score final ISQ 2018	99%	65%
Équité	Score final ISQ 2018	100%	83%
Engagement des parents	Efficacité	92%	92%
Équité	Efficacité	85%	88%
Efficiace	Efficacité	96%	58%
Efficacité	Efficiace	96%	47%

Les relations entre l'efficacité et l'efficiace sont singulières comme le montrent les deux dernières lignes du tableau : la probabilité de présence est forte et identique : 96%. Par contre, la probabilité de direction est de 56% dans le sens « Efficiace → Efficacité » alors qu'elle est de 45% dans le sens « Efficacité → Efficiace ». Ce deuxième sens est une dépendance fonctionnelle : le calcul de l'efficiace dépend de l'efficacité qui est rapportée au niveau des moyens mis en œuvre.

4.2 Principaux résultats issus du DAG

Le diagramme (Figure 3) ainsi que les résultats du Tableau 1 mettent en avant les résultats suivants :

Figure 3 : Graphe acyclique orienté (DAG), avec les dépendances fonctionnelles



- De toutes les dépendances fonctionnelles, l'équité – définie comme la capacité d'un système scolaire à compenser l'impact de l'origine sociale sur les performances scolaires – semble avoir le rôle le plus important pour l'ensemble des pays, sachant que cette importance doit certainement varier d'un pays à l'autre. Cette relation fonctionnelle est la seule présente à 100% pour expliquer l'ISQ. L'équité a aussi une relation causale forte pour expliquer l'efficacité. Cela pourrait signifier que plus un système éducatif est équitable, plus il est efficace. En d'autres termes, accorder par exemple une attention particulière aux élèves qui possèdent un moindre vocabulaire que les autres (ce qui nuit à leur bonne compréhension du cours), ou qui rencontrent des difficultés d'attention ou de concentration ou qui enfin ne maîtrisent pas les codes scolaires, contribuerait, suivant les pays, à accroître l'efficacité des systèmes éducatifs et leur qualité.
- L'engagement des parents a un statut particulier : sa dépendance fonctionnelle vers l'ISQ 2018 est relativement faible. Comme nous l'avons mis en avant dans une publication précédente (Gerard, Hugonnier et Varin, à paraître), l'introduction de ce facteur modifie peu le score ISQ, c'est-à-dire le niveau de qualité global des systèmes éducatifs des pays de l'OCDE dans leur ensemble. L'hypothèse explicative que nous avons avancée était liée à l'âge des élèves concernés par les études PISA : 15 ans. À cet âge, en effet, les élèves semblent moins attentifs et sensibles aux conseils de leurs parents. Certains travaux (Fan et Chen, 2001 ; Hoover-Dempsey, Battiato, Walker, Reed, DeJong and Jones, 2001 ; Organisation de coopération et de développement économiques [OCDE], 2019) indiquent par ailleurs que l'implication élevée des parents de jeunes de cet âge peut avoir un effet négatif sur leurs résultats, contrairement à ce qui semble être le cas avec des enfants au début des apprentissages scolaires.

- En revanche, l'établissement du DAG indique un effet causal important de l'engagement des parents sur l'efficacité. C'est même la relation la plus forte des dépendances probabilistes. Cette particularité pourrait s'expliquer par le fait que, pour la plupart des parents, ce qui compte avant tout, ce sont les notes de leurs enfants à l'école (ce que reflète en partie le facteur efficacité), et ce d'autant plus que c'est quasiment la seule information dont ils disposent quant aux performances de leurs enfants (à moins de prendre rendez-vous avec le professeur principal). Ce constat ne peut que nous inviter à continuer à prendre en compte la dimension « engagement des parents » dans nos analyses futures, même si son impact sur l'ISQ est relativement faible.
- L'efficacité – c'est-à-dire la capacité d'un système scolaire à permettre aux élèves d'avoir de bons résultats scolaires – est un facteur qui influe non seulement directement sur le score final de l'ISQ, mais également à la suite de relations causales exercées sur elle-même par l'engagement des parents, l'efficacité et l'équité en tant que dépendances probabilistes. En cela, l'efficacité est un facteur singulier.
- Ce n'est pas un hasard si ce facteur est souvent considéré comme la principale caractéristique de la performance, voire de la qualité d'un système éducatif : cela n'aurait en effet aucun sens d'avancer qu'un système éducatif est de qualité s'il ne permet pas d'atteindre ses objectifs en termes d'apprentissage des élèves. Nos recherches sur le DAG établissent de plus que cette place importante de l'efficacité s'explique par d'autres facteurs : l'équité, l'engagement des parents et l'efficacité.
- Comme nous l'avons signalé lors de la présentation du Tableau 1, la dépendance fonctionnelle entre l'efficacité et l'efficacité est singulière. Fortement présente dans les différentes itérations de calcul (96%). Dans 58% des itérations, la relation indique que c'est l'efficacité qui influence l'efficacité, alors que la direction inverse est présente à 47%. Ce constat est étonnant, car par construction, l'efficacité dépend de l'efficacité (qui est donc une dépendance fonctionnelle) : l'efficacité est en effet le rapport entre l'efficacité et les moyens mis en œuvre pour l'atteindre. Par définition, à moyens égaux, plus l'efficacité est importante, plus l'efficacité est élevée. Or, les résultats du DAG privilégient l'hypothèse causale inverse : plus l'efficacité est forte, plus l'efficacité l'est aussi. Cette influence exercée par l'efficacité sur l'efficacité pourrait tenir au fait qu'au cas où l'efficacité est insuffisante, les pressions politiques sont vives pour que des mesures soient prises afin d'accroître le niveau d'efficacité.
- Le fait que les six critères ont une dépendance fonctionnelle directe et élevée sur l'ISQ 2018 est logique. Néanmoins, ce constat est important, car il montre comment l'utilisation des réseaux bayésiens causaux, et singulièrement celui du DAG, permet d'identifier des

relations causales non perceptibles par des techniques statistiques classiques. Dans un autre article, nous avons montré, grâce au coefficient tau de Kendall, que la présence ou non d'un seul critère ne modifiait pas de manière significative le score ISQ des pays (Gerard, Hugonnier et Varin, à paraître). Grâce au DAG, nous pouvons nuancer ce constat : les dépendances fonctionnelles entre chaque critère et le niveau de qualité des systèmes éducatifs peuvent être quantifiées. Les dépendances fonctionnelles les plus fortes concernent l'équité et l'engagement des élèves, ce qui en soi est un résultat très important en matière de politique éducative.

5. Conclusion

Par cette recherche et l'utilisation de l'outil méthodologique qu'est le DAG, nous avons mis en évidence des liens de causalité qui jusqu'à présent étaient impossibles à identifier. Les autorités politiques décisionnelles gagneraient à prendre connaissance de ces travaux afin de pouvoir agir de manière positive sur la qualité des systèmes éducatifs. En effet, il ressort de manière claire que l'équité et dans une moindre mesure l'engagement des parents et l'efficacité sont trois leviers efficaces sur lesquels le monde politique pourrait agir concrètement afin d'essayer d'augmenter la qualité des systèmes éducatifs.

Il reste que, comme déjà signalé, l'analyse faite ici porte sur l'ensemble des pays de l'OCDE. De sorte qu'établir qu'œuvrer pour accroître l'équité, l'engagement des parents et l'efficacité peuvent avoir un impact important sur la qualité des systèmes éducatifs des pays de l'OCDE, ne permet pas de conclure que cela vaut pour tous les pays pris individuellement. Dans tous nos travaux (Gerard, Hugonnier et Varin, 2017, 2018, à paraître), nous avons répété que ce qui importe est que chaque pays, grâce à l'ISQ, puisse analyser sa situation personnelle et prendre les décisions qui lui conviennent. À cet égard, dans ces autres articles, les résultats, critère par critère, pour chaque pays sont indiqués. C'est sur cette base que des pays peuvent décider ou non de prendre des mesures qui les concernent. L'approche du DAG permet en plus d'indiquer qu'il y a – globalement – des influences plus fortes que d'autres. Cela aide à comprendre la dynamique globale, mais ne doit surtout pas mener à former de mêmes préconisations politiques pour tous les pays.

Sans conteste, les outils d'analyse utilisés dans le présent article permettent de mettre en évidence leur grande utilité pour être enfin en mesure de discerner les liens de causalité entre diverses variables, ce qui ouvre de grandes perspectives pour les sciences sociales en général et plus particulièrement pour les recherches sur l'indicateur synthétique mesurant la qualité des systèmes éducatifs des pays de l'OCDE (ISQ). En effet, passer des corrélations partielles à des relations causales constitue pour les chercheurs une opportunité pour mieux approfondir les

connaissances des phénomènes tant sociaux, économiques que scientifiques, même si les réseaux bayésiens doivent répondre à de strictes conditions avant de pouvoir être utilisés comme réseaux bayésiens causaux.

6. Références

- Ben Hassen, H., Masmoudi, A. and Rebai, A. (2008). Causal inference in bio-molecular pathways using a bayesian network approach and an implicit method. *Journal of Theoretical Biology*, 253(4) :717 – 724.
- Cartwright, N. (1994). *Nature's Capacities and Their Measurement*. Oxford University Press.
- Chickering, D., Geiger, D. et Heckerman, D. (1995). Learning bayesian networks : Search methods and experimental results. *Proceedings of Fifth Conference on Artificial Intelligence and Statistics*, pages 112–128.
- Chickering, D. et Heckerman, D. (1996). Efficient Approximation for the Marginal Likelihood of Incomplete Data given a Bayesian Network. *UAI'96*, pages 158–168. Morgan Kaufmann.
- Fan, X., Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review* 13, 1–22.
- Friedman, N., Goldszmidt, M. & Wyner, A. (1999). Data analysis with bayesian networks: A bootstrap approach. *UAI'99 : Proceedings of the 15th annual conference on uncertainty in artificial intelligence*, 196-205.
- Gámez, J.A., Mateo, J.L. & Puerta, J.M. (2011). Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Min Knowl Disc* 22, 106–148.
- Gerard, F.-M., Hugonnier, B. & Varin, S. (2017). La qualité des systèmes éducatifs des pays de l'OCDE enfin mesurée, in B. Hugonnier & G. Serrano (dir.). *Réconcilier la République et son école*. Paris : Éditions du Cerf, pp. 61-73.
- Gerard, F.-M., Hugonnier, B. & Varin, S. (2018). Mesure de la qualité des systèmes éducatifs des pays de l'OCDE, in ADMEE-Europe, *L'évaluation en éducation et en formation face aux transformations des sociétés contemporaines, Actes du colloque*, Esch-sur-Alzette : Université de Luxembourg, pp. 131-143.
- Gerard, F.-M., Hugonnier, B. & Varin, S. (à paraître). Indicateur synthétique de la qualité des systèmes éducatifs des pays de l'OCDE : comparaison des résultats 2015 et 2018,
- Hernán, M.A., Robins, J.M. (2018). *Causal Inference*. Chapman and Hall/CRC.
- Hoover-Dempsey, K. V., Battiato, A. C., Walker, J. M. T., Reed, R. P., DeJong, J. M., & Jones, K. P. (2001). Parental involvement in homework. *Educational Psychologist*, 36(3), 195–209.
- Imbens, G.W., Rubin, D.B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge university press.
- Kalisch, M., Mächler, M., Colombo, D., Maathuis, MH., Bühlmann, P. (2012). “Causal Inference Using Graphical Models with the R Package pcalg.” *Journal of Statistical Software*, 47(11), 1–26.
- Lauritzen, S. (1996). *Graphical Models*. Oxford University Press, Oxford.
- Maathuis, MH., Colombo, D., Kalisch, M., Bühlmann, P. (2010). Predicting Causal Effects in Large-Scale Systems from Observational Data. *Nature Methods*, 7, 261–278.
- Maathuis, MH., Kalisch, M., Bühlmann, P. (2009). Estimating High-Dimensional Intervention Effects from Observational Data. *The Annals of Statistics*, 37, 3133–3164.
- Morgan, S.L., Winship, C. (2014). *Counterfactuals and Causal Inference*. Cambridge university press, 2nd édition.

- Murphy, K. (2007). *How to use the Bayes Net Toolbox*. Available at : <https://web.archive.org/web/20140626142205/http://bnt.googlecode.com/svn/trunk/docs/usage.html>
- Naïm, P., Wuillemin, P-H., Leray, P., Pourret, O. et Becker, A. (2007). *Réseaux bayésiens*. Eyrolles, Paris, 3^e édition.
- Naïm, P., Pourret, O and Marcot, B. (2008). Dirichlet process gaussian mixture models : Choice of the base distribution. *Bayesian Networks : A Practical Guide to Applications*, Wiley.
- Organisation de coopération et de développement économiques (2019). *Résultats du PISA 2018 (Volume I) : Savoirs et savoir-faire des élèves*. Éditions OCDE. <https://doi.org/10.1787/ec30bc50-fr>
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4) :669–688.
- Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge university press.
- Pearl, J. (2003). Statistics and causal inference : A review. *Test*, 12(2) :281–345.
- Pearl, J. (2009). *Causality : Models, Reasoning, and Inference*. Cambridge University Press, 2nd edition.
- Peters, J., Janzing, D. (2017). *Elements of Causal Inference - Foundations and Learning Algorithms*. MIT Press.
- R Core Team (2021). *R : A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Scutari, M. (2017). Bayesian Network Constraint-Based Structure Learning Algorithms: Parallel and Optimized Implementations in the bnlearn R Package. *Journal of Statistical Software*, 77(2), 1-20. doi:10.18637/jss.v077.i02.
- Spirtes, P., Meek, C. and Richardson, T. (1999). An algorithm for causal inference in the presence of latent variables and selection bias. In *Computation, Causation, and Discovery*, 211–252. AAAI Press, Menlo Park, CA
- Spirtes, P., Glymour, C., Scheines, R. (2000). *Causation, Prediction, and Search*. *Adaptive Computation and Machine Learning*, 2nd edition. MIT Press, Cambridge.
- Talbot, D. (2012). *Introduction à une approche graphique d'inférence causale*. Association des statisticiennes et statisticiens du Québec. Disponible sur : <https://www.association-assq.qc.ca/2012/02/16/introduction-a-une-approche-graphique-dinferece-causale/>
- Vanderweele, T.J. (2015). *Explanation in Casual Inference : Methods for Mediation and Interaction*. Oxford University Press.